

会议文献出版特点及其资源建设策略研究

□ 赖院根 / 中国科学技术信息研究所 北京 100038
李俊莉 / 中国科学技术信息研究所 北京 100038
/ 河南警察学院 郑州 450000

摘要:会议文献是重要的信息资源,但当前图书馆对其管理比较分散,资源建设缺乏系统性。对CPCI收录会议数据展开分析以了解其出版特点,结果发现:80%以上的会议文献在会议召开后两年内出版;会议文献在“工程技术”和“医药卫生”领域数量较多;连续性会议名称非常不规范等。在此基础上,对会议文献资源建设现状、难点进行了分析,并提出了有针对性的资源建设策略。

关键词:会议文献,学术会议,资源建设,CPCI

DOI: 10.3772/j.issn.1673—2286.2013.05.001

1 引言

学术会议文献是在各种学术会议上发表的文献的统称,其出版形式比较多样,包括会议录、论文集、图书、期刊特辑、科技报告等多种形式^[1]。会议文献含有大量的最新科技信息,是了解世界科技发展动向、水平和最新成就的主要渠道,是一种具有很高的参考价值的信息资源^[2]。但在资源建设与服务上,当前会议文献管理比较分散,用户获取仍然相对困难^[3]。

随着数字出版、数字存取技术的兴起和发展,近年来会议文献的受重视程度有所提高。在出版数量上,在1980至2008年期间,IEEE出版的期刊论文数量增加了4倍,会议论文数量则增加了近20倍^[4]。在论文使用上,会议文献已经成为计算机学科引文的主要文献类型^[5];

ACM出版论文的参考文献中有40%为会议文献^[6]。与此同时,国内外学术会议召开数量明显增多,并呈现专业化、市场化、精品化的发展趋势^[7]。有理由相信,会议文献的影响力还将得到进一步的提升。面对当前发展现状,从满足用户需求的角度出发,图书馆很有必要加强会议文献资源建设工作。那么,会议文献资源建设的难点有哪些?应该采取什么样的策略来改变当前现状?现有文献鲜有讨论这些问题,本文研究因此而产生。

本文拟首先对CPCI数据库(Conference Proceedings Citation Index,原ISTP)收录的学术会议数据展开定量分析,目的是了解会议文献出版特点。在此基础上,对会议文献资源建设现状、难点进行讨论并提出有针对性的应对策略。本文的研究逻辑是只有在了解和把握

会议文献出版特点的基础上,讨论资源建设方案才能言之有据,才能更好地推动图书馆资源建设工作向定量化和精细化发展。

2 分析数据

CPCI是内容最全面、覆盖学科最广泛的学术会议论文数据库,由美国科学情报研究所(ISI)编辑出版,被视为世界范围内最权威的学术会议文献索引工具。鉴于其在学术领域的被认可程度,本文选择该数据库作为会议文献出版特点研究的数据源。

从ISI网站(www.webofknowledge.com)下载CPCI收录的学术会议信息并逐条核实,共得到记录128,795条,下载时间为2012年4月11日。数据字段包括会议名称、会议地点、会议日期、卷、

期、出版时间、ISSN、ISBN、学科类别检索词等。

对下载数据进行整理,具体步骤包括:(1)信息抽取,例如从“会议地点”中抽取“会议召开国别”、从“会议日期”中抽取“会议召开年份”、从“出版时间”中抽取“出版年份”等。抽取过程中对“字段内容为空”或“数据错误”等现象,统一以“-1”标识其字段内容。(2)按“会议名称”、“会议召开国别”和“会议召开年份”去重,剩余记录个数为126,937。以这些记录为基础,建立了学术会议分析数据仓库,其中出版年份在1990-2010年期间的会议文献数量见图1。

从图1上看,CPCI对1995年及以前出版的会议文献收录较少,每年数量只有几百册;自1998年起,收录数量上升到年均9000册这一量级。由于世界上召开的学术会议数量在1995年至1998年期间不大可能发生如此大的波动,图中收录文献数量的巨大变化只能解释为ISI对会议文献重视程度的明显提高,这从侧面反映出本文研究的意义与价值。Web of Science在世界范围内拥有庞大的用户群,CPCI数据库的建设与完善有利于提高会议文献的可获得性和利用率。

关于分析数据有几点需要说明:(1)学术会议可能是每年召开

的“连续性会议”,也可能是只召开一次的专题性会议,两者在统计时很容易造成混淆。本文对连续性会议分开统计,不同时间召开的连续性会议被统计成多个会议;(2)规模较大的学术会议可能会有多册纸质出版物,在严格意义上此类会议出版物的计量单位应该为“套”而不是“册”。由于CPCI以“会议名称”作为收录关键词,因此没必要再区分“套”与“册”之间的差异。为方便表述,本文统一以“册”作为会议文献的计量单位。简单地说,本文分析的126,937条记录,可以理解成CPCI收录的126,937个分开统计的学术会议或126,937册会议文献。

3 会议文献出版特点

3.1 出版时滞

会议文献涉及两个时间,一个是会议召开时间,另一个是文献出版时间。通常情况下,出版时间要晚于会议召开时间。为了解出版时滞分布,以文献出版年份减去会议召开年份并汇总,结果见表1(以年为单位进行统计主要是由于许多记录只有年份数据,同时也是为了计算上的简便,但由此不可避免地会带来一些误差)。从表中数据看,约20%的会议文献在会议召开当年就能刊出,80%以上会在会议召开2年内出版。

3.2 类别检索词分布

Web of Science类别检索词是ISI分配给每一记录的传统检索词。对CPCI会议文献记录的类别检索词进行汇总,发现它们分别属于270个“Web of Science

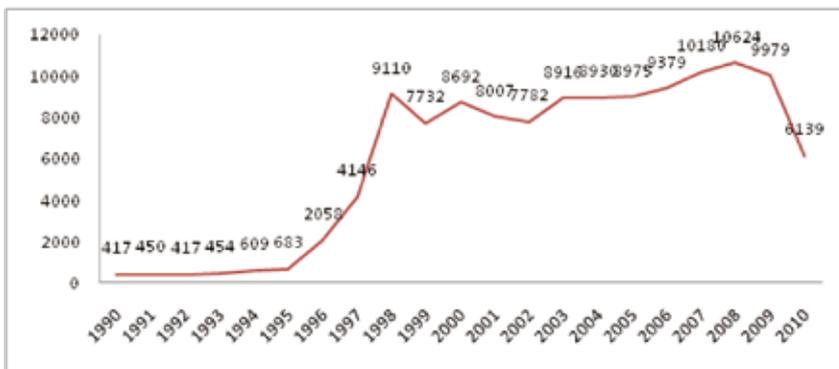


图1 会议文献出版年份分布

表1 出版时滞分析

滞后时间	数量	比例
0年	27482	21.7%
1年	45473	35.8%
2年	29957	23.6%
3年	13463	10.6%
4年	5844	4.6%
5年及以上	2829	2.2%
无效数据	1889	1.5%

类别”（字段内容为空的92条记录不参与统计），其中超过1500册的类别检索词有16个，见表2。由于类别划分较细，每个类别检索词对应的会议文献数量并不是很多，最多的是“工程，电气（ENGINEERING, ELECTRICAL & ELECTRONIC）”，共3131册。表中有5个都与“计算机科学”相关，累计和为11410册，接近总数的10%。从前16类别分布上看，会议

文献在“工程技术”（包括表2中的“工程、电气”，“计算机科学”等）、“医药卫生”（包括表2中的“外科”、“肿瘤学”、“心脏与心血管系统”等）等领域较多，这一点与前人的研究结果比较一致^[8]。

3.3 出版形式

会议文献出版形式多样，包括论文集图书、期刊专辑等多种类

型。根据Web of Science网站上的介绍，CPCI收录的会议文献部分来自科技期刊，部分来自会议论文集。其中，物理和工程领域的会议论文主要以论文集图书的方式出版，生物医药领域和生命学科则多选择期刊形式出版。

为了解CPCI收录数据的出版形式，对以期刊形式出版的文献数量进行统计，结果见表3。根据图1中的文献数量变化，表3还将

表2 会议文献学科分布 (>1500册)

序号	类别检索词	数量
1	工程，电气（ENGINEERING, ELECTRICAL & ELECTRONIC）	3131
2	外科（SURGERY）	3054
3	计算机科学，人工智能（COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE）	2971
4	经济（ECONOMICS）	2795
5	计算机科学，信息系统（COMPUTER SCIENCE, INFORMATION SYSTEMS）	2718
6	计算机科学，理论与方法（COMPUTER SCIENCE, THEORY & METHODS）	2422
7	肿瘤学（ONCOLOGY）	2204
8	材料科学，多学科（MATERIALS SCIENCE, MULTIDISCIPLINARY）	2054
9	心脏与心血管系统（CARDIAC & CARDIOVASCULAR SYSTEMS）	1962
10	生物化学与分子生物学（BIOCHEMISTRY & MOLECULAR BIOLOGY）	1961
11	临床神经病学（CLINICAL NEUROLOGY）	1933
12	计算机科学，软件工程（COMPUTER SCIENCE, SOFTWARE ENGINEERING）	1770
13	药学，普通与内科（MEDICINE, GENERAL & INTERNAL）	1698
14	免疫学（IMMUNOLOGY）	1613
15	光学（OPTICS）	1570
16	计算机科学，硬件与架构（COMPUTER SCIENCE, HARDWARE & ARCHITECTURE）	1529

表3 出版形式统计

序号	1998年之前		1998年及以后		整体	
	数量	比例	数量	比例	数量	比例
会议文献总数	9649	-	117288	-	126937	-
期刊形式出版	3778	39.2%	83937	71.6%	87715	69.1%
非期刊形式出版	5871	60.8%	33351	28.4%	39222	30.9%

所有年份划分为“1998年之前”和“1998年及以后”两阶段进行统计以方便比较。由于下载数据中没有“出版形式”这一字段,本文结合科技期刊特点进行判定,条件为记录中有准确的ISSN号或有完整的卷号和期号。这样统计可能会存在一些计算误差,实际计算值可能要大于表中比例。

从表3可以看出,CPCI收录的会议文献近70%以期刊形式出版;1998年之前出版的会议文献,该比例不到40%。如此高的期刊形式出版比例主要跟ISI自身特点及其收录标准有关。ISI拥有三大期刊引文数据库,从中抽取与学术会议相关的论文数据非常便利。需要说明的是,由于无法得知世界范围内究竟有多少比例的会议文献以期刊形式出版,也缺乏“出版形式”与“会议质量”相关关系的实证研究,表中的数据只能反映CPCI数据库的收录特点,而不能视为会议文献整体的出版趋势或特点。

3.4 会议连续性

如前文所述,学术会议可以划分为连续性会议和一次性会议。一般来说,连续性会议会使用年份/届次不同、其他关键词相同的名称,因此可以根据会议名称或其缩写来判断是否属于同一个系列会议。考虑到连续性会议很少间隔4年才召开,同时受人力和时间限制,本文选择2005-2009年期间出版的会议文献数据进行分析,记录个数为39422条。分析思路是先对会议名称进行规范,然后在汇总的基础上得到连续性会议数量与比例。

会议名称规范工作包括两个步骤,首先是编写程序对会议名称进行自动处理;其次是人工对记录逐条进行核实。整理结果发现会议名称非常杂乱,不规范和拼写错误非常严重,例如:(1)会议名称中年份、届次位置不定,居前(如“1997 Information-and-System-Society Conference of IEICE”)、居中(如“AAAI 2001 Spring Symposium on Answer Set Programming”)和居后(如“ACM Multimedia Conference 2007”)

的情况都有,还存在年份简写或用序号表示届次的现象(如“SAS 98 Conference”、“XVIth EURO Conference”);(2)多个会议联合召开,比如“25th ACM SIGMOD International Conference on Management of Data/Symposium on Principles of Database Systems”;(3)属于同一个会议但历年名称有些差别,比如“of”与“for”互换、关键词多一个或少一个“s”、个别关键词次序颠倒等。数据杂乱导致的直接结果就是工作量明显增大,这也是仅对部分CPCI数据进行整理的主要原因。

对规范后的数据进行汇总,结果见表4。表中“连续次数”列表示会议名称汇总时的出现次数,其中“一次有届次”指会议名称只出现一次但名称中含有届次序号(如2nd、3rd等)或含有“annual”字样。本文认为“一次有届次”会议也属于连续性会议,只是受收录标准、出版等因素的影响,使得其在2005-2009年期间只被CPCI收录一次。

从表4上看,2005-2009年的39422册会议文献共涉及29281个会议,其中连续性会议个数为16905个,约占总数的57.7%;非连续性会议(即表中“一次没届次”)比例为42.3%。在相关会议文献上,约70%为连续性会议的出版物。需要说明的是,虽然名称规范主要由人工完成,但仍然存在许多难以判定的情况;表中认定为非连续性会议的名称中,部分包含季节性词汇(如“spring”、“summer”等),整理时未对其进行处理。

表4 会议连续性分布(出版年份为2005-2009年)

序号	学术会议		相关会议文献	
	数量	比例	数量	比例
五次及以上	557	1.9%	2966	7.5%
四次	595	2.0%	2380	6.0%
三次	1215	4.1%	3645	9.2%
两次	3517	12.0%	7034	17.8%
一次有届次	11021	37.6%	11021	28.0%
一次没届次	12376	42.3%	12376	31.4%
合计	29281	-	39422	-
连续性会议	16905	57.7%	27046	68.6%
非连续性会议	12376	42.3%	12376	31.4%

4 资源建设现状与策略

根据上节对CPCI收录数据的

分析结果,下面对我国会议文献资源建设现状与难点进行分析,并提出有针对性的应对策略。

4.1 现状与难点

数字环境下,会议文献的学术价值被重新评价,这一点从图1中CPCI收录数量的变化上体现得非常明显,但当前其建设状况并不系统,体现在:(1)资源分散收藏。这种分散既体现在大部分图书馆都只收藏部分会议文献资料上,也体现在同一个图书馆拥有的会议文献分散在图书、期刊、缩微平片等不同载体中。(2)加工揭示缺乏统一规范。许多图书馆都建有相对完善的期刊书目控制系统,但很少设置专门机构对会议文献进行管理,一般都是图书形式出版的会议文献按图书处理,期刊形式出版的按期刊处理。(3)缺乏统一、便利的检索平台。虽然国内外已经有一些会议文献专题数据库,包括前文分析的CPCI数据库,但由于覆盖率或数据加工揭示等方面的原因,不同需求仍需采取不同的获取策略^[9]。

仔细分析造成当前现状的原因,笔者认为主要跟两方面有关。首先,长期以来,会议文献的学术价值被认为要低于科技期刊^[10],原因包括被引用比例偏低^[11]、学科分布侧重于工程与应用领域^[12]等。其次,受会议文献自身特点的影响,虽然科技期刊和会议文献都属于学术交流媒介,都能反映科技投入产出,但两者在许多方面存在明显差异。在出版方式上,科技期刊公开发行,学术层次、每年出版期次和论文数量相对稳定;会议文献属于灰色文献,只有部分公开出版,学术水平差异大。其中,当前建设

难点主要受后者的影响,归纳起来有以下几点:

(1)会议文献质量良莠不齐,缺乏广泛认可的遴选标准。科技期刊出版相对稳定,并有国际上广为接受的期刊影响因子作为评价指标,使其资源遴选工作能够比较有序和系统地进行。但对会议文献而言,目前既无一个类似于乌利希期刊指南的权威性出版信息数据库,也没有普遍认可的评价体系。在缺乏理论和出版信息支撑的情况下,会议文献资源建设工作只能依靠经验完成,这难免造成杂乱、分散的局面。

(2)学术会议召开规律性不强,很难进行跟踪或追溯。虽然学术会议发展呈现信息化、精品化等趋势,也的确存在一些影响力较高的连续性会议,但总的来说影响学术会议召开的因素很多,甚至受国际政治、经济环境变化方面的影响。并且,会议主办方可能变换出版形式或出版商,出版时滞又长短不一(参见表1),这就进一步加大了会议文献出版信息的搜集工作难度。

(3)会议文献著录项相对特殊,数据规范工作量大。与科技期刊相比,会议文献除了常见的著录项外,还有会议名称、届次、会议地点等字段,目前对会议文献如何进行加工揭示缺乏统一的规范说明。此外,与期刊、图书等一般使用论文题名进行检索不同,会议文献经常使用会议名称这一字段进行检索,这样就有必要对连续性会议进行规范加工以利于用户进行浏览和获取。但从前文的叙述可以知晓,要实现这一目标工作量很大。

4.2 建设策略

针对当前现状,比较理想的解

决方案是建立完整的会议文献出版信息数据库和科学的会议文献评估方法与遴选标准。但从目前可获知的信息上看,要实现这两个目标还有相当长的一段距离。为有效改变当前现状,图书馆在开展会议文献资源建设时可选择如下策略:

(1)加强与学协会的联系,充分利用其机构网站信息。当前国内图书馆一般通过代理公司查找、订购会议文献,当代理公司数据更新不及时就容易影响订到率^[13]。学协会是学术会议的主要组织者之一,是许多重要学术会议的主办机构^[14]。加强与学协会的联系,在绕过代理公司这一中间环节后,有利于获取第一手的会议召开信息,更好地把握“信息源”;同时,学协会对其主办的会议更为了解,通过它们可以对会议规模、连续性等有更清晰的认识,从而提高资源遴选科学性。

(2)分学科开展会议文献资源建设,加强对CPCI数据的分析处理。从前文的分析可以看出,会议文献数量及其出版形式在学科之间存在很大的差别(参见表2),因此在工作中有必要区别对待。CPCI收录的会议文献70%以期刊形式出版(参见表3),大部分为连续性会议的出版物(参见表4),并且通过Web of Science平台能够得到会议论文的被引用数据。可见CPCI数据具有很好的参考作用,既可以通过对它的分析来获得连续性会议召开及其出版商信息,也可以用于对会议等级进行评价和划分^[15],这对完善会议文献资源建设工作无疑有着重要价值。

(3)加强对已有资源的分析利用,拓展服务方式,重视用户推介。满足用户需求是文献管理机构工作的出发点和归宿,也是资源建设的

根本目的。许多机构已经拥有相当数量的会议文献数据,只是分散于不同媒介之中。如能很好地进行规范整理并向用户提供服务,既能充分利用已有资源,还能更好地了解用户需求及其特征,必要时也可以通过用户推荐的方式来建立会议文献资源列表。

5 结语

数字环境下,会议文献的受重

视程度明显提高,但当前资源建设缺乏系统性,用户获取仍然比较困难。为了解当前会议文献的出版特点,本文选择CPCI收录数据从出版时滞、学科分布、出版形式和会议连续性等方面进行了分析。在此基础上,分析了会议文献资源建设难点并提出了有针对性的应对策略。

本文的研究尚存在许多局限性。首先,CPCI数据的分析结果只能反映CPCI的收录特点,未必是所

有会议文献的出版特点,这也是无论选择哪个数据源都必然存在的问题;其次,当前与会议文献研究相关的研究很少,尚未看到会议文献出版特点方面的研究成果。作为一次尝试,本文虽然分析出一些结果,但对出版特点和建设策略之间关系的讨论还不够深入。这些方面也是下一步研究需要逐步加强的地方。

参考文献

- [1] 赖茂生,徐克敏,等.科技文献检索(第二版)[M].北京:北京大学出版社,1994:30-32.
- [2] 袁军鹏.科学计量学高级教程[M].北京:科学技术文献出版社,2010:68-70.
- [3] 赵美娣.会议文献的检索与获取[J].情报理论与实践,2011,34(8):84-86.
- [4] SHIRAKAWA N, FURUKAWA T, NOMURA M, et al. Global Competition and Technological Transition in Electrical, Electronic, Information and Communication Engineering: Quantitative Analysis of Periodicals and Conference Proceedings of the IEEE [J]. Scientometrics, 2012, 91(3): 895-910.
- [5] BAR-ILAN J. Web of Science with the Conference Proceedings Citation Indexes: the Case of Computer Science [J]. Scientometrics, 2010, 83(3): 809-824.
- [6] WAINER J, OLIVEIRA H P, ANIDO R. Patterns of Bibliographic References in the ACM Published Papers [J]. Information Processing and Management, 2011, 47(1): 135-142.
- [7] 刘兴平.学术会议的兴起与发展[J].科技导报,2010(6):19-24.
- [8] WOLFGANG G, BALÁZS S, ANDRÁS S, et al. Proceedings Literature as Additional Data Source for Bibliometric Analysis [J]. Scientometrics, 2006, 68(3): 457-473.
- [9] 任平.网络环境下会议文献的新特点与获取之道[J].图书情报工作,2007,51(1):108-110,103.
- [10] GODIN B. Measuring Knowledge Flows between Countries: the Use of Scientific Meeting Data [J]. Scientometrics, 1998, 42(3): 313-323.
- [11] CYNTHIA L, VINCENT L, ÉRIC A. Conference Proceedings as a Source of Scientific Information: A Bibliometric Analysis [J]. Journal of the American Society for Information Science and Technology, 2008, 59(11): 1776-1784.
- [12] WOLEK F W, GRIFFITH B. C. Policy and Informal Communication in Applied Science and Technology [J]. Science Studies, 1974(4): 411-420.
- [13] 顾德南,袁晓鹏,马占营.试论网络环境下外文会议文献的采访[J].冶金信息导刊,2009(2):43-46,50.
- [14] 中国科协“学会学术交流活动状况及期望专题调研组”.学术交流是学会凝聚科技工作者的基础——学会学术交流活动状况及期望调查报告[J].学会,2005(7):16-25.
- [15] 孙乐民,谢永强,施燕斌.用ISTP统计评估国际学术会议的等级[J].高校图书馆工作,2008(5):44-45,49.

作者简介

赖根根(1977-),男,博士,副研究员,研究方向:知识组织、数据挖掘和用户研究。E-mail: lai yg@istic.ac.cn

Study on Publishing Characteristics of Conference Literature and Strategies for Collection Development

Lai Yuangen / Institute of Scientific & Technical Information of China, Beijing, 100038

Li Junli / Institute of Scientific & Technical Information of China, Beijing, 100038

/ Henan Police College, Zhengzhou, 450000

Abstract: The conference literature is one of the important information resources, while its organization in libraries is dispersed and the collection development is short of consistency currently. This paper makes a discussion on CPCI data to learn its publishing characteristics. The results shows: firstly, more than 80 percent of conference literature is published within two years after the date of academic conferences. Secondly, the number of conference literature in engineering and medicine fields is larger. Thirdly, names of successive conferences are very irregular. Furthermore, current situation and difficulties are discussed, and some targeted development strategies of conference literature have been proposed.

Keywords: Conference literature, Academic conference, Collection development, CPCI

(收稿日期: 2013-03-25)